Appendices - In-Silico Evaluation of YCT-529

Capstone Project

Contents

Appendix A: Thesis Statement

In this study, I argue that physiologically based pharmacokinetic (PBPK) modeling provides a reliable framework for predicting the absorption, distribution, metabolism, and excretion (ADME) properties of YCT-529, a novel non-hormonal male contraceptive. By leveraging PK-Sim and ADMETLab simulations, I demonstrate that YCT-529 exhibits significant dose- and population-dependent pharmacokinetics, with East Asians showing delayed clearance and Black Americans exhibiting lower systemic exposure due to variations in CYP3A4, CYP2C8, and UGT1A1 activity [36, 11]. These computational predictions highlight the need for population-specific dosing regimens to ensure optimal therapeutic efficacy while minimizing variability-induced risks. Additionally, this study emphasizes the influence of dietary intake on systemic exposure, underscoring the need to explore the practical and statistical significance of these differences for dosing optimization. Integrating computational pharmacokinetics into early drug development provides a cost-effective, scalable strategy for refining dosing protocols before experimental validation, addressing critical gaps in male contraceptive research and advancing personalized pharmacotherapy.

Appendix B: Medium Article

For a more intuitive and accessible explanation of the pharmacokinetics of YCT-529, including its mechanism as a male contraceptive, please refer to the Medium article: The Science of a Male Birth Control Pill: A Deep Dive into YCT-529's Pharmacokinetics.

Appendix C: Extended Discussion on Computational Models

SMILES String Generation

The SMILES (Simplified Molecular Input Line Entry System) representation is foundational in computational chemistry for encoding chemical structures into a compact, alphanumeric string that facilitates computational processing. The process of generating the SMILES string for YCT-529 required meticulous attention to molecular detail, as even small inaccuracies can disrupt downstream applications in cheminformatics and pharmacokinetics. This section provides an extended discussion on the technical, functional, and validation aspects of SMILES string generation.

The JSME Molecule Editor, integrated into the Cheminfo.org platform, serves as a key tool for this process. It is a robust JavaScript-based software that eliminates the need for standalone molecular drawing tools by allowing users to construct molecular

structures directly within a web browser. This tool is an evolution of the original JME applet, developed to ensure compatibility with modern web technologies while preserving its robust chemical drawing capabilities [4]. The JSME editor supports various chemical conventions, including bond types, stereochemistry, and functional group representations, ensuring the accuracy of the generated SMILES string.

For YCT-529, the construction of the molecular structure involved replicating key features such as its carboxylic acid group and aromatic systems, which are essential for its activity as a non-hormonal male contraceptive. The molecular structure was referenced from the PubChem database [20], which provided a reliable and standardized depiction of the compound. By carefully inputting these features into the JSME editor, I ensured fidelity to the actual molecular configuration of YCT-529.

Technical Aspects

The SMILES string generation process involves translating the visual representation of a molecule into a linear, textual format.

`Cc5ccc(C3=CC(C)(C)Oc4ccc(c2ccc(c1ccc(C(=O)O)cc1)[nH]2)cc34)cc5`.

The string was produced for YCT-529, encapsulating its molecular framework and functional groups. Each character in the string corresponds to a specific atom, bond type, or structural feature. For instance:

- **C** represents carbon atoms.
- = denotes a double bond.
- **Brackets** [**nH**] specify a nitrogen atom with an attached hydrogen, crucial for representing aromatic systems.

The hierarchical and recursive structure of SMILES strings allows for efficient storage and computational analysis of complex molecules. Moreover, SMILES strings can be converted back into 2D or 3D molecular structures using cheminformatics tools, confirming the integrity of the encoded data.

Validation of the generated SMILES string is a critical step to ensure that it accurately represents the intended molecule. I utilized ChemDraw, a widely recognized tool in molecular design, to verify the structural fidelity of the SMILES string. This software enables back-conversion of SMILES into molecular diagrams, ensuring that the chemical structure matches its intended design. Any discrepancies at this stage would have necessitated re-evaluation and correction within the JSME editor.

Role of SMILES in Computational Workflows

The role of SMILES in computational workflows extends far beyond being a simple molecular representation. As the input format for tools like ADMETLab 3.0 and PK-Sim, the

SMILES string serves as the foundation for predicting physicochemical properties, ADME behavior, and pharmacokinetic profiles. For instance, ADMETLab uses SMILES to calculate metrics such as LogP, LogD, and solubility, which are critical for modeling drug behavior in biological systems [15]. PK-Sim further uses this data to simulate absorption, distribution, metabolism, and excretion processes, which are discussed in subsequent sections of this appendix.

While the SMILES format is widely adopted, it has limitations. For example, it does not inherently encode 3D spatial data such as stereochemistry or conformational flexibility, which may influence a molecule's pharmacokinetics. These aspects must be addressed using additional cheminformatics tools or enhanced SMILES variants like InChI (International Chemical Identifier). Moreover, the accuracy of the SMILES string is contingent upon the precision of the initial molecular drawing, emphasizing the need for careful construction and verification.

The generation and validation of the SMILES string for YCT-529 represent a critical first step in its computational analysis. By ensuring the accuracy and fidelity of the SMILES string, I established a robust foundation for the advanced cheminformatics and pharmacokinetic simulations detailed in this study. This extended discussion highlights the technical and functional considerations essential for researchers seeking to adopt similar workflows in drug development.

ADMETLab 3.0

While the main text provides a thorough overview of how ADMETLab 3.0 was utilized to analyze the physicochemical properties, medicinal chemistry rules, ADME metrics, and toxicity profiles of YCT-529, further information is available for readers seeking a deeper understanding of the tool and its capabilities. The ADMETLab 3.0 Documentation (https://admetlab3.scbdd.com/documentation/#/new) provides detailed insights into the platform's operational framework, including the methodologies used to calculate and predict various endpoints. Additionally, the Endpoints Explanation (https://admetlab3.scbdd.com/explanation/#/) expands on the meanings, significance, and relevance of the calculated properties such as LogP, LogS, TPSA, and ADME-specific metrics. These resources are invaluable for understanding how ADMETLab 3.0 integrates cheminformatics and computational modeling techniques to predict compound behavior and ensure that researchers can interpret the results with precision.

For example, properties like LogP (partition coefficient) and TPSA (topological polar surface area) play critical roles in drug absorption and distribution, as they influence membrane permeability and solubility. The documentation explains the mathematical and computational foundations of these metrics, aiding researchers in comprehending their significance. Similarly, the toxicity endpoints, such as hepatotoxicity or genotoxic-

ity, are backed by robust computational models that the documentation elaborates on, offering clarity for those delving into the complex interplay of a compound's properties and its pharmacological implications.

PK-Sim Setup

Initially, I considered using BioGears, a C++-based human physiology engine known for its extensive modeling capabilities and open-source framework [5]. While BioGears offers a robust simulation environment, I found it challenging to install and run due to my limited familiarity with C++ and the incompatibility with my ARM64-based macOS system. Even after reaching out to a BioGears developer, we could not resolve the installation issues. In contrast, when I switched to PK-Sim, I discovered a more user-friendly, GUI-driven platform that, although not native to macOS, I could easily run on a borrowed Windows machine. PK-Sim's comprehensive documentation, physiologically realistic compartment models, and straightforward handling of inter-individual variability aligned closely with my research objectives. Moreover, the tool's cost-free availability and established track record in PBPK simulations made it a practical and methodologically sound alternative to more complex platforms like BioGears.

PK-Sim is a versatile platform for physiologically based pharmacokinetic/pharmacodynamic (PBPK/PD) modeling. Readers interested in understanding the capabilities of this tool can refer to the detailed PK-Sim Documentation. This resource comprehensively explains the software's functionalities, including the creation of individuals and populations, compound definition, and administration protocol configuration. For an in-depth understanding of its theoretical foundations, the Modeling Concepts section discusses compartmental and non-compartmental analysis techniques central to PK-Sim's modeling framework.

These resources expand significantly on the topics covered in the main text, particularly on PK-Sim's robust handling of interindividual variability, population-based simulations, and dynamic parameter adjustments. For example, the documentation explains how population variability is modeled through Monte Carlo simulations and provides guidance on customizing enzyme expression levels and organ-specific physiologies to align with experimental or clinical data. Furthermore, the modeling concepts delve into how PK-Sim handles complex drug behaviors, such as nonlinear kinetics or interactions with metabolizing enzymes like CYP450s, offering clarity on the rationale behind the model design choices. A video demonstrating how to navigate my YCT-529 PK-Sim model is available here.

Appendix D: Data Pre-Processing for Pharmacokinetic Analysis

After running my simulations in PK-Sim, I used its built-in graphing tools to visualize the pharmacokinetic results. However, these tools were limited in flexibility, preventing key modifications such as aggregating results across multiple simulations, customizing graph aesthetics (e.g., removing gridlines or adding detailed titles), and making comparative overlays of YCT-529 against MCC or Atazanavir. To overcome these constraints, I exported the raw simulation data and performed all data processing and visualization in Python.

PK-Sim generates two key datasets for each simulation: the Results file and the PK-Analyses file.

1. Results File: This dataset provides time-resolved concentration data, recording drug distribution at multiple physiological compartments. The key columns include:

- IndividualId: Identifies each simulated subject.
- Time [min]: Simulation time in minutes.
- Organism|PeripheralVenousBlood|YCT-529|Plasma (Peripheral Venous Blood) [µmol/L]: Drug concentration in venous plasma.
- Organism|VenousBlood|Plasma|YCT-529|Concentration in container [µmol/L]: Another measure of venous plasma concentration.
- Organism|Lumen|YCT-529|Fraction of oral drug mass absorbed into mucosa: Fraction of the administered dose absorbed into the gastrointestinal mucosa.

This file enabled me to generate plasma concentration-time curves for YCT-529 across all races and doses.

2. PK-Analyses File: This dataset summarizes key pharmacokinetic parameters extracted from the simulation. The key columns include:

- IndividualId: Identifies each simulated subject.
- QuantityPath: Indicates the physiological location of the measurement.
- Parameter: Reports pharmacokinetic metrics such as C_{max} (maximum plasma concentration).
- Value: Numerical value of the given pharmacokinetic parameter.
- Unit: Measurement unit (e.g., µmol/L).

This file was essential for analyzing peak plasma concentrations across races and doses.

Data Pre-Processing and Aggregation Before importing these datasets into Python for analysis, I modified them by adding two critical columns: Race and Dosage (mg). The updated structure of the results file was as follows:

Race	Dosage (mg)	IndividualId	Time [min]	Plasma (Peripheral Venous Blood) [µmol/L]	Venous Plasma [µmol/L]	Fraction Absorbed
Black American	15	0	0	0.000	0.000	0.000
Black American	15	0	3	0.0000863	0.000238	0.0213
Black American	15	0	6	0.000746	0.001301	0.0491
...	...	...	...	...	...	...

By including Race and Dosage (mg), I was able to aggregate all simulation data into a unified dataset. This enabled meaningful comparisons across populations and doses, facilitating the creation of comprehensive visualizations such as:

- Venous plasma concentration vs. time plots for individual populations and doses.
- C_{max} distributions across different racial groups.
- Dietary impact analysis on YCT-529 pharmacokinetics.
- Comparative overlays of YCT-529, MCC, and Atazanavir.

Once aggregated, I imported the pre-processed dataset into Python and used pandas, matplotlib, and seaborn to analyze the data and generate all plots. This pre-processing step was crucial in ensuring that my dataset was structured optimally for efficient and reproducible pharmacokinetic analysis. The Python scripts in this repository detail the full data-processing workflow, including statistical calculations and visualizations.

By transitioning from PK-Sim's built-in tools to Python-based data processing, I was able to refine my analysis, enhance visualization capabilities, and ensure greater flexibility in statistical interpretation. A video explaining the full data pre-processing workflow, including Python-based analysis, can be found here.

Appendix E: Python-Based Data Analysis and Visualization

To analyze and visualize the pharmacokinetic data generated from PK-Sim simulations, I utilized Python, specifically leveraging the pandas, matplotlib, and seaborn libraries. PK-Sim provides built-in visualization tools, but they lack the flexibility needed for advanced data processing, comparative analysis, and customized visualizations. Therefore, I exported the simulation data and pre-processed it in Python to enable more refined analyses.

All Python scripts, along with the dataset and simulation files, are available in the GitHub repository: GitHub Repository for YCT-529 PBPK Modeling.

Data Availability and Execution

The dataset required for the analysis is available on Google Drive: Capstone Research Data.

To run the analysis, follow these steps:

- Download the dataset from Google Drive.
- Upload the dataset to Google Colab or your local Jupyter Notebook environment.
- Open the `PBPK_data_analysis.ipynb` file and execute the code step-by-step.
- Alternatively, you can run the Python script `pbpk_data_analysis.py` in a terminal or IDE.

Google Colab is the recommended execution platform, as it allows direct file uploads and execution without local dependencies.

Python Code for Data Analysis and Visualization

Below is an excerpt of the Python code used to process and visualize the simulation data:

```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Define file paths
file_paths = {
    "Black American": "BAM Aggregated Sim Protocol-Results.csv",
    "White American": "WAM Aggregated Sim Protocol-Results.csv",
    "East Asian": "EastAsianMen Aggregated Sim Protocol-Results.csv",
    "European": "EuroMen Aggregated Sim Protocol-Results.csv",
    "MCC Placebo": "MCC Placebo Aggregated Sim Protocol-Results.csv"
```

```
}

# Molecular weight for unit conversion
MW_YCT_529 = 435.5  # g/mol

# Load and process each race's data
race_data = {}
time_points = None

for race, file in file_paths.items():
    df = pd.read_csv(file)

    # Convert time from minutes to hours
    df["Time (hours)"] = df["Time [min]"] / 60

    # Extract venous plasma concentration and convert from  mol /L to g /L
    plasma_col = "Organism|VenousBlood|Plasma|YCT-529|Concentration in container [$\mu$mol/l]"
    df["Plasma Concentration ($\mu$g/L)"] = df[plasma_col] * MW_YCT_529

    if time_points is None:
        time_points = df["Time (hours)"].unique()

    # Compute average for each time point
    avg_concentration = df.groupby("Time (hours)")["Plasma Concentration ($\mu$g/L)"].mean()
    race_data[race] = avg_concentration

# Convert race data to DataFrame
df_plot = pd.DataFrame(race_data)

# Plot data
plt.figure(figsize=(8, 5))
for race in race_data.keys():
    plt.plot(df_plot.index, df_plot[race], label=race, linewidth=2)

plt.xlabel("Time (hours)")
plt.ylabel(r"Plasma Concentration ($\mu$g/L)")
plt.title("YCT-529 Pharmacokinetics Across Populations")
plt.legend()
plt.grid(False)
plt.show()
```

Listing 1: Processing and Plotting YCT-529 Pharmacokinetics

To quantify the impact of dietary conditions on YCT-529 pharmacokinetics, I also performed both statistical and practical significance tests on C_{max} values obtained from

PBPK simulation outputs. A one-way ANOVA confirmed statistically significant differences between groups ($F = 793.45$, $p < 0.001$), and Tukey's HSD post-hoc test revealed significant differences between the fasted state versus both the light meal ($p < 0.001$, $d = 1.38$) and high-fat breakfast ($p < 0.001$, $d = 1.61$). No significant difference was observed between light meal and high-fat conditions ($p = 0.842$, $d = 0.22$). These analyses confirm that dietary intake meaningfully alters YCT-529 absorption.

The Python script below demonstrates the ANOVA and Cohen's d calculations:

```python
from scipy.stats import f_oneway
from statsmodels.stats.multicomp import pairwise_tukeyhsd
import numpy as np

# Grouping by dietary condition
c_max_df = df.groupby(['IndividualId', 'Dietary Condition'])['Plasma Concentration ( g /L)'].max().reset_index()
fasted = c_max_df[c_max_df['Dietary Condition'] == 'Fasted']['Plasma Concentration ( g /L)']
light = c_max_df[c_max_df['Dietary Condition'] == 'Light Meal']['Plasma Concentration ( g /L)']
fatty = c_max_df[c_max_df['Dietary Condition'] == 'High Fat Breakfast']['Plasma Concentration ( g /L)']

# One-way ANOVA
f_stat, p_value = f_oneway(fasted, light, fatty)

# Cohens d
def cohens_d(group1, group2):
    pooled_std = np.sqrt(((len(group1) - 1) * np.var(group1) +
                          (len(group2) - 1) * np.var(group2)) /
                         (len(group1) + len(group2) - 2))
    return (np.mean(group1) - np.mean(group2)) / pooled_std
```

Listing 2: Statistical and practical significance testing of Cmax values

The rest of the code, including statistical calculations and advanced visualizations, can be found in the GitHub repository: here.

Key Takeaways

- Python-based data processing allowed for aggregating and comparing YCT-529 pharmacokinetics across populations and doses.
- Google Colab provides an accessible platform for executing the scripts without local setup.
- The complete dataset must be downloaded and uploaded into the notebook for analysis.

By transitioning from PK-Sim's built-in tools to Python-based analysis, I was able to generate detailed insights, enhance visualization clarity, and enable flexible data manipulation for further pharmacokinetic assessments. A video explaining the data preprocessing and Python-based analysis can be found here.

Appendix ϕ: ADMETLab Results

YCT-529 ADMETLab Results

1. Physicochemical Property

Property	Value	Comment
Molecular Weight	435.18	Contain hydrogen atoms. Optimal:100~600
Volume	470.451	Van der Waals volume
Density	0.925	Density = MW / Volume
nHA	4.0	Number of hydrogen bond acceptors. Optimal:0~12
nHD	2.0	Number of hydrogen bond donors. Optimal:0~7
nRot	4.0	Number of rotatable bonds. Optimal:0~11
nRing	5.0	Number of rings. Optimal:0~6
MaxRing	10.0	Number of atoms in the biggest ring. Optimal:0~18
nHet	4.0	Number of heteroatoms. Optimal:1~15
fChar	0.0	Formal charge. Optimal:-4 ~4
nRig	29.0	Number of rigid bonds. Optimal:0~30
Flexibility	0.138	Flexibility = nRot /nRig
Stereo Centers	0.0	Stereo Centers. Optimal: ≤ 2
TPSA	62.32	Topological Polar Surface Area. Optimal:0~140
logS	-5.761	The logarithm of aqueous solubility value.
logP	5.706	The logarithm of the n-octanol/water distribution coefficients at pH=7.4.
logD	3.356	The logarithm of the n-octanol/water distribution coefficient.
pka (Acid)	6.671	Acid-base dissociation constant (pKa) value represents the strength of a drug molecule's acidity or basicity.
pka (Base)	2.89	Acid-base dissociation constant (pKa) value represents the strength of a drug molecule's acidity or basicity.
Melting point	223.983	The predicted melting point of a compound is expressed in degrees Celsius (°C). Melting points below 25°C are classified as liquids, while melting points above 25°C are classified as solids.
Boiling point	426.848	The predicted melting point of a compound is expressed in degrees Celsius (°C). A normal boiling point below 25°C is categorized as a gas.

2. Medicinal Chemistry

Property	Value	Decision	Comment

QED	0.364	🔴	■ A measure of drug-likeness based on the concept of desirability; ■ Attractive: > 0.67; ■ unattractive: 0.49~0.67; ■ too complex: < 0.34
GASA	0.0	🟢	■ ES: Easy to synthesize; HS: Hard to synthesize; ■ The output value represents the probability of being difficult to synthesize, ranging from 0 to 1.
Synth	2.0	🟢	■ Synthetic accessibility score is designed to estimate ease of synthesis of drug-like molecules. ■ SAscore ≥ 6, difficult to synthesize; SAscore <6, easy to synthesize
Fsp3	0.138	🔴	■ The number of sp3 hybridized carbons / total carbon count, correlating with melting point and solubility. ■ Fsp^3 ≥0.42 is considered a suitable value.
MCE-18	59.758	🟢	■ MCE-18 stands for medicinal chemistry evolution. ■ MCE-18≥45 is considered a suitable value.
NPscore	0.253	-	■ Natural product-likeness score. ■ This score is typically in the range from -5 to 5. ■ The higher the score is, the higher the probability is that the molecule is a NP.
Lipinski Rule	0.0	🟢	■ MW ≤ 500; logP ≤ 5; Hacc ≤ 10; Hdon ≤ 5 ■ If two properties are out of range, a poor absorption or permeability is possible, one is acceptable.
Pfizer Rule	1.0	🔴	■ logP > 3; TPSA < 75 ■ Compounds with a high log P (>3) and low TPSA (<75) are likely to be toxic.
GSK Rule	1.0	🔴	■ MW ≤ 400; logP ≤ 4 ■ Compounds satisfying the GSK rule may have a more favorable ADMET profile
Golden Triangle	0.0	🟢	■ 200 ≤ MW ≤ 500; -2 ≤ logD ≤ 5 ■ Compounds satisfying the Golden Triangle rule may have a more favorable ADMET profile.
PAINS	0 alerts	-	frequent hitters, Alpha-screen artifacts and reactive compound 480 substructures (J Med Chem 201053:2719-40)
ALARM NMR	0 alerts	-	Thiol reactive compounds.
BMS	0 alerts	-	undesirable, reactive compounds 176 substructures (J Chem Inf Model 200646:1060-8)
Chelator Rule	0 alerts	-	Chelating compounds.
Colloidal aggregators	1.0	-	■ Category 0: non-colloidal aggregators; ■ Category 1: colloidal aggregators. ■ The output value is the probability of being colloidal aggregators, within the range of 0 to 1.

FLuc inhibitors	0.951	●	■ Category 0: non-fLuc inhibitors; ■ Category 1: fLuc inhibitors. ■ The output value is the probability of being fLuc inhibitors, within the range of 0 to 1.
Blue fluorescence	0.843	●	■ Category 0: non-blue fluorescence; ■ Category 1: blue fluorescence. ■ The output value is the probability of being blue fluorescence, within the range of 0 to 1.
Green fluorescence	0.941	●	■ Category 0: non-green fluorescence; ■ Category 1: green fluorescence. ■ The output value is the probability of being green fluorescence, within the range of 0 to 1.
Reactive compounds	0.212	●	■ Category 0: non-reactive compound; ■ Category 1: reactive compound. ■ The output value is the probability of being reactive compound, within the range of 0 to 1.
Promiscuous compounds	0.224	●	■ Category 0: non-promiscuous compound; ■ Category 1: promiscuous compound. ■ The output value is the probability of being promiscuous compound, within the range of 0 to 1.

3. Absorption

Property	Value	Decision	Comment
Caco-2 Permeability	-5.12	●	Optimal: higher than -5.15 Log unit
MDCK Permeability	-4.8	●	■ low permeability: < 2 × 10-6 cm/s ■ medium permeability: 2-20 × 10^{-6} cm/s ■ high passive permeability: > 20 × 10^{-6} cm/s
PAMPA	0.96	●	■ The experimental data for Peff was logarithmically transformed (logPeff). ■ Molecules with log Peff values below 2.0 were classified as low-permeability (Category 0), while those with log Peff values exceeding 2.5 were classified as high-permeability (Category 1).
Pgp-inhibitor	0.478	●	■ Category 1: Inhibitor; ■ Category 0: Non-inhibitor; ■ The output value is the probability of being Pgp-inhibitor
Pgp-substrate	0.023	●	■ Category 1: substrate; ■ Category 0: Non-substrate; ■ The output value is the probability of being Pgp-substrate
HIA	0.001	●	■ Human Intestinal Absorption ■ Category 1: HIA+(HIA < 30%); ■ Category 0: HIA-(HIA >= 30%); ■ The output value is the probability of being HIA+

$F_{20\%}$	0.132	●	■ 20% Bioavailability ■ Category 1: F 20% + (bioavailability < 20%); ■ Category 0: F 20% - (bioavailability ≥ 20%); ■ The output value is the probability of being F 20% +
$F_{30\%}$	0.089	●	■ 30% Bioavailability ■ Category 1: F 30% + (bioavailability < 30%); ■ Category 0: F 30% - (bioavailability ≥ 30%); ■ The output value is the probability of being F 30% +
$F_{50\%}$	0.798	●	■ 50% Bioavailability ■ Category 1: F 50% + (bioavailability < 50%); ■ Category 0: F 50% - (bioavailability ≥ 50%); ■ The output value is the probability of being F 50% +

4. Distribution

Property	Value	Decision	Comment
PPB	98.817	●	■ Plasma Protein Binding Optimal: < 90%. ■ Drugs with high protein-bound may have a low therapeutic index.
VDss	-0.414	●	■ Volume Distribution ■ Optimal: 0.04-20L/kg
BBB	0.005	●	■ Blood-Brain Barrier Penetration ■ Category 1: BBB+; Category 0: BBB-; ■ The output value is the probability of being BBB+
Fu	0.537	●	■ The fraction unbound in plasms ■ Low: <5%; Middle: 5~20%; High: > 20%
OATP1B1 inhibitor	0.999	●	■ Category 0: Non-inhibitor; Category 1: inhibitor. ■ The output value is the probability of being inhibitor, within the range of 0 to 1.
OATP1B3 inhibitor	0.997	●	■ Category 0: Non-inhibitor; Category 1: inhibitor. ■ The output value is the probability of being inhibitor, within the range of 0 to 1.
BCRP inhibitor	0.013	●	■ Category 0: Non-inhibitor; Category 1: inhibitor. ■ The output value is the probability of being inhibitor, within the range of 0 to 1.
MRP1 inhibitor	0.993	●	■ Category 0: Non-inhibitor; Category 1: inhibitor. ■ The output value is the probability of being inhibitor, within the range of 0 to 1.

5. Metabolism

Property	Value	Decision	Comment
CYP1A2 inhibitor	0.698	●	■ Category 1: Inhibitor; Category 0: Non-inhibitor; ■ The output value is the probability of being inhibitor.

CYP1A2 substrate	0.0	●	■ Category 1: Substrate; Category 0: Non-substrate; ■ The output value is the probability of being substrate.
CYP2C19 inhibitor	1.0	●	■ Category 1: Inhibitor; Category 0: Non-inhibitor; ■ The output value is the probability of being inhibitor.
CYP2C19 substrate	0.0	●	■ Category 1: Substrate; Category 0: Non-substrate; ■ The output value is the probability of being substrate.
CYP2C9 inhibitor	0.139	●	■ Category 1: Inhibitor; Category 0: Non-inhibitor; ■ The output value is the probability of being inhibitor.
CYP2C9 substrate	0.005	●	■ Category 1: Substrate; Category 0: Non-substrate; ■ The output value is the probability of being substrate.
CYP2D6 inhibitor	0.001	●	■ Category 1: Inhibitor; Category 0: Non-inhibitor; ■ The output value is the probability of being inhibitor.
CYP2D6 substrate	0.0	●	■ Category 1: Substrate; Category 0: Non-substrate; ■ The output value is the probability of being substrate.
CYP3A4 inhibitor	0.0	●	■ Category 1: Inhibitor; Category 0: Non-inhibitor; ■ The output value is the probability of being inhibitor.
CYP3A4 substrate	0.001	●	■ Category 1: Substrate; Category 0: Non-substrate; ■ The output value is the probability of being substrate.
CYP2B6 inhibitor	0.998	●	■ Category 1: Inhibitor; Category 0: Non-inhibitor; ■ The output value is the probability of being inhibitor.
CYP2B6 substrate	0.21	●	■ Category 1: Substrate; Category 0: Non-substrate; ■ The output value is the probability of being substrate.
CYP2C8 inhibitor	1.0	●	■ Category 1: Inhibitor; Category 0: Non-inhibitor; ■ The output value is the probability of being inhibitor.
HLM Stability	0.121	●	■ human liver microsomal (HLM) stability ■ Category 0: stable+ (HLM > 30 min); Category 1: unstable- (HLM ≤ 30 min). The output value is the probability of human liver microsomal instability, where a value closer to 1 indicates a higher likelihood of instability. The range is between 0 and 1.

6. Excretion

Property	Value	Decision	Comment

CL_{plasma}	0.384	●	■ The unit of predicted CLplasma penetration is ml/min/kg. >15 ml/min/kg: high clearance; 5-15 ml/min/kg: moderate clearance; < 5 ml/min/kg: low clearance.
$T_{1/2}$	1.286	●	■ The unit of predicted T1/2 is hours. ■ ultra-short half-life drugs: 1/2 < 1 hour; short half-life drugs: T1/2 between 1-4 hours; intermediate short half-life drugs: T1/2 between 4-8 hours; long half-life drugs: T1/2 > 8 hours.

7. Toxicity

Property	Value	Decision	Comment
hERG Blockers	0.825	●	■ Molecules with IC50 ≤10μM or ≥50% inhibition at 10 μM were classified as hERG+ (Category 1), ■ while molecules with IC50 >10μM or < 50% inhibition at 10μM were classified as hERG - (Category 0). ■ The output value is the probability of being hERG+, within the range of 0 to 1.
hERG Blockers (10um)	0.333	●	■ Molecules with IC50 ≤10 μM are classified as hERG+ (Category 1), ■ and molecules with IC50 > 10μM are classified as hERG- (Category 0). ■ The output value is the probability of being hERG+, within the range of 0 to 1.
DILI	0.998	●	■ Drug Induced Liver Injury. ■ Category 1: drugs with a high risk of DILI; ■ Category 0: drugs with no risk of DILI. ■ The output value is the probability of being toxic.
AMES Mutagenicity	0.909	●	■ AMES Toxicity ■ Category 1: Ames positive(+); ■ Category 0: Ames negative(-); ■ The output value is the probability of being toxic.
Rat Oral Acute Toxicity	0.708	●	■ Rat Oral Acute Toxicity. ■ Category 0: low-toxicity, > 500 mg/kg; ■ Category 1: high-toxicity; < 500 mg/kg. ■ The output value is the probability of being toxic, within the range of 0 to 1.
FDAMDD	0.955	●	■ FDA Maximum (Recommended) Daily Dose. ■ Category 1: FDAMDD (+); ■ Category 0: FDAMDD (-); The output value is the probability of being positive.
Skin Sensitization	0.059	●	■ Category 1: Sensitizer; ■ Category 0: Non-sensitizer. ■ The output value is the probability of being toxic, within the range of 0 to 1.
Carcinogenicity	0.772	●	■ Category 1: carcinogens; ■ Category 0: non-carcinogens; ■ The output value is the probability of being toxic.

Eye Corrosion	0.0	●	■ Eye Corrosion ■ Category 1: corrosives; Category 0: noncorrosives; ■ The output value is the probability of being corrosives.
Eye Irritation	0.163	●	■ Eye Irritation ■ Category 1: irritants; Category 0: nonirritants; ■ The output value is the probability of being irritants.
Respiratory	0.944	●	■ Category 1: respiratory toxicants; ■ Category 0: non-respiratory toxicants. ■ The output value is the probability of being toxic, within the range of 0 to 1.
Human Hepatotoxicity	0.888	●	■ Human Hepatotoxicity ■ Category 1: H-HT positive(+); ■ Category 0: H-HT negative(-); ■ The output value is the probability of being toxic.
Drug-induced Nephrotoxicity	0.993	●	■ Category 0: non-nephrotoxic (-); ■ Category 1: nephrotoxic (+). ■ The output value is the probability of being nephrotoxic (+), within the range of 0 to 1.
Ototoxicity	0.942	●	■ Category 0: non-ototoxicity (-); ■ Category 1: ototoxicity (+). ■ The output value is the probability of being ototoxicity (+), within the range of 0 to 1.
Hematotoxicity	0.702	●	■ Category 0: non-hematotoxicity (-); ■ Category 1: hematotoxicity (+). ■ The output value is the probability of being hematotoxicity (+), within the range of 0 to 1.
Genotoxicity	0.983	●	■ Category 0: non-Genotoxicity (-); ■ Category 1: Genotoxicity (+). ■ The output value is the probability of being ototoxicity (+), within the range of 0 to 1.
RPMI-8226 Immunitoxicity	0.229	●	■ Category 0: non-cytotoxicity (-); ■ Category 1: cytotoxicity (+). ■ The output value is the probability of being ototoxicity (+), within the range of 0 to 1.
A549 Cytotoxicity	0.562	●	■ Category 0: non-cytotoxicity (-); ■ Category 1: cytotoxicity (+). ■ The output value is the probability of being ototoxicity (+), within the range of 0 to 1.
Hek293 Cytotoxicity	0.827	●	■ Category 0: non-cytotoxicity (-); ■ Category 1: cytotoxicity (+). ■ The output value is the probability of being ototoxicity (+), within the range of 0 to 1.
Drug-induced Neurotoxicity	0.942	●	■ Category 0: non-neurotoxic (-); ■ Category 1: neurotoxic (+). ■ The output value is the probability of being neurotoxic (+), within the range of 0 to 1.

8. Environmental toxicity

Property	Value	Comment
Bioconcentration Factors	1.451	■ Bioconcentration factors are used for considering secondary poisoning potential and assessing risks to human health via the food chain. ■ The unit is –log10[(mg/L)/(1000*MW)]
IGC_{50}	4.614	■ Tetrahymena pyriformis 50 percent growth inhibition concentration. ■ The unit is –log10[(mg/L)/(1000*MW)]
$LC_{50}FM$	5.941	■ 96-hour fathead minnow 50 percent lethal concentration. ■ The unit is –log10[(mg/L)/(1000*MW)]
$LC_{50}DM$	5.989	■ 48-hour daphnia magna 50 percent lethal concentration. ■ The unit is –log10[(mg/L)/(1000*MW)]

9. Tox21 pathway

Property	Value	Decision	Comment
NR-AhR	0.872	🔴	■ Aryl hydrocarbon receptor ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
NR-AR	0.066	🟢	■ Androgen receptor ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
NR-AR-LBD	0.003	🟢	■ Androgen receptor ligand-binding domain ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
NR-Aromatase	0.014	🟢	■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
NR-ER	0.489	🟠	■ Estrogen receptor ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
NR-ER-LBD	0.002	🟢	■ Estrogen receptor ligand-binding domain ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
NR-PPAR-gamma	0.02	🟢	■ Peroxisome proliferator-activated receptor gamma ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
SR-ARE	0.7	🟠	■ Antioxidant response element ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.

SR-ATAD5	0.001	●	■ ATPase family AAA domain-containing protein 5 ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
SR-HSE	0.36	●	■ Heat shock factor response element ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
SR-MMP	0.655	●	■ Mitochondrial membrane potential ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
SR-p53	0.01	●	■ p53, a tumor suppressor protein ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.

10. Toxicophore Rules

Property	Value	Comment
Acute Toxicity Rule	0	■ 20 substructures; ■ acute toxicity during oral administration
Genotoxic Carcinogenicity Rule	0	■ 117 substructures; ■ carcinogenicity or mutagenicity
NonGenotoxic Carcinogenicity Rule	0	■ 23 substructures; ■ carcinogenicity through nongenotoxic mechanisms
Skin Sensitization Rule	0	■ 155 substructures; ■ skin irritation
Aquatic Toxicity Rule	0	■ 99 substructures; ■ toxicity to liquid(water)
NonBiodegradable Rule	0	■ 19 substructures; ■ non-biodegradable
SureChEMBL Rule	0	■ 164 substructures; ■ MedChem unfriendly status
FAF-Drugs4 Rule	1 alerts	154 toxic substructures from FAF-Drug4

D-Glucose ADMETLab Results

1. Physicochemical Property

Property	Value	Comment
Molecular Weight	180.06	Contain hydrogen atoms. Optimal:100~600
Volume	156.517	Van der Waals volume
Density	1.15	Density = MW / Volume
nHA	6.0	Number of hydrogen bond acceptors. Optimal:0~12
nHD	5.0	Number of hydrogen bond donors. Optimal:0~7
nRot	1.0	Number of rotatable bonds. Optimal:0~11
nRing	1.0	Number of rings. Optimal:0~6
MaxRing	6.0	Number of atoms in the biggest ring. Optimal:0~18
nHet	6.0	Number of heteroatoms. Optimal:1~15
fChar	0.0	Formal charge. Optimal:-4 ~4
nRig	6.0	Number of rigid bonds. Optimal:0~30
Flexibility	0.167	Flexibility = nRot /nRig
Stereo Centers	5.0	Stereo Centers. Optimal: ≤ 2
TPSA	110.38	Topological Polar Surface Area. Optimal:0~140
logS	0.339	The logarithm of aqueous solubility value.
logP	-2.229	The logarithm of the n-octanol/water distribution coefficients at pH=7.4.
logD	-2.092	The logarithm of the n-octanol/water distribution coefficient.
pka (Acid)	10.761	Acid-base dissociation constant (pKa) value represents the strength of a drug molecule's acidity or basicity.
pka (Base)	5.139	Acid-base dissociation constant (pKa) value represents the strength of a drug molecule's acidity or basicity.
Melting point	88.769	The predicted melting point of a compound is expressed in degrees Celsius (°C). Melting points below 25°C are classified as liquids, while melting points above 25°C are classified as solids.
Boiling point	272.225	The predicted melting point of a compound is expressed in degrees Celsius (°C). A normal boiling point below 25°C is categorized as a gas.

2. Medicinal Chemistry

Property	Value	Decision	Comment

QED	0.29	🔴	■ A measure of drug-likeness based on the concept of desirability; ■ Attractive: > 0.67; ■ unattractive: 0.49~0.67; ■ too complex: < 0.34
GASA	0.0	🟢	■ ES: Easy to synthesize; HS: Hard to synthesize; ■ The output value represents the probability of being difficult to synthesize, ranging from 0 to 1.
Synth	3.0	🟢	■ Synthetic accessibility score is designed to estimate ease of synthesis of drug-like molecules. ■ SAscore ≥ 6, difficult to synthesize; SAscore <6, easy to synthesize
Fsp3	1.0	🟢	■ The number of sp3 hybridized carbons / total carbon count, correlating with melting point and solubility. ■ Fsp^3 ≥0.42 is considered a suitable value.
MCE-18	22.667	🔴	■ MCE-18 stands for medicinal chemistry evolution. ■ MCE-18≥45 is considered a suitable value.
NPscore	2.627	-	■ Natural product-likeness score. ■ This score is typically in the range from -5 to 5. ■ The higher the score is, the higher the probability is that the molecule is a NP.
Lipinski Rule	0.0	🟢	■ MW ≤ 500; logP ≤ 5; Hacc ≤ 10; Hdon ≤ 5 ■ If two properties are out of range, a poor absorption or permeability is possible, one is acceptable.
Pfizer Rule	0.0	🟢	■ logP > 3; TPSA < 75 ■ Compounds with a high log P (>3) and low TPSA (<75) are likely to be toxic.
GSK Rule	0.0	🟢	■ MW ≤ 400; logP ≤ 4 ■ Compounds satisfying the GSK rule may have a more favorable ADMET profile
Golden Triangle	1.0	🔴	■ 200 ≤ MW ≤ 500; -2 ≤ logD ≤ 5 ■ Compounds satisfying the Golden Triangle rule may have a more favorable ADMET profile.
PAINS	0 alerts	-	frequent hitters, Alpha-screen artifacts and reactive compound 480 substructures (J Med Chem 201053:2719-40)
ALARM NMR	0 alerts	-	Thiol reactive compounds.
BMS	0 alerts	-	undesirable, reactive compounds 176 substructures (J Chem Inf Model 200646:1060-8)
Chelator Rule	0 alerts	-	Chelating compounds.
Colloidal aggregators	0.016	-	■ Category 0: non-colloidal aggregators; ■ Category 1: colloidal aggregators. ■ The output value is the probability of being colloidal aggregators, within the range of 0 to 1.

FLuc inhibitors	0.0	●	■ Category 0: non-fLuc inhibitors; ■ Category 1: fLuc inhibitors. ■ The output value is the probability of being fLuc inhibitors, within the range of 0 to 1.
Blue fluorescence	0.084	●	■ Category 0: non-blue fluorescence; ■ Category 1: blue fluorescence. ■ The output value is the probability of being blue fluorescence, within the range of 0 to 1.
Green fluorescence	0.0	●	■ Category 0: non-green fluorescence; ■ Category 1: green fluorescence. ■ The output value is the probability of being green fluorescence, within the range of 0 to 1.
Reactive compounds	0.492	●	■ Category 0: non-reactive compound; ■ Category 1: reactive compound. ■ The output value is the probability of being reactive compound, within the range of 0 to 1.
Promiscuous compounds	0.068	●	■ Category 0: non-promiscuous compound; ■ Category 1: promiscuous compound. ■ The output value is the probability of being promiscuous compound, within the range of 0 to 1.

3. Absorption

Property	Value	Decision	Comment
Caco-2 Permeability	-6.408	●	Optimal: higher than -5.15 Log unit
MDCK Permeability	-4.907	●	■ low permeability: < 2 × 10-6 cm/s ■ medium permeability: 2-20 × 10^{-6} cm/s ■ high passive permeability: > 20 × 10^{-6} cm/s
PAMPA	0.988	●	■ The experimental data for Peff was logarithmically transformed (logPeff). ■ Molecules with log Peff values below 2.0 were classified as low-permeability (Category 0), while those with log Peff values exceeding 2.5 were classified as high-permeability (Category 1).
Pgp-inhibitor	0.0	●	■ Category 1: Inhibitor; ■ Category 0: Non-inhibitor; ■ The output value is the probability of being Pgp-inhibitor
Pgp-substrate	0.626	●	■ Category 1: substrate; ■ Category 0: Non-substrate; ■ The output value is the probability of being Pgp-substrate
HIA	0.958	●	■ Human Intestinal Absorption ■ Category 1: HIA+(HIA < 30%); ■ Category 0: HIA-(HIA >= 30%); ■ The output value is the probability of being HIA+

$F_{20\%}$	0.46	🟠	■ 20% Bioavailability ■ Category 1: F 20% + (bioavailability < 20%); ■ Category 0: F 20% - (bioavailability ≥ 20%); ■ The output value is the probability of being F 20% +
$F_{30\%}$	0.948	🔴	■ 30% Bioavailability ■ Category 1: F 30% + (bioavailability < 30%); ■ Category 0: F 30% - (bioavailability ≥ 30%); ■ The output value is the probability of being F 30% +
$F_{50\%}$	0.849	🔴	■ 50% Bioavailability ■ Category 1: F 50% + (bioavailability < 50%); ■ Category 0: F 50% - (bioavailability ≥ 50%); ■ The output value is the probability of being F 50% +

4. Distribution

Property	Value	Decision	Comment
PPB	29.711	🟢	■ Plasma Protein Binding Optimal: < 90%. ■ Drugs with high protein-bound may have a low therapeutic index.
VDss	-0.511	🔴	■ Volume Distribution ■ Optimal: 0.04-20L/kg
BBB	0.581	🟠	■ Blood-Brain Barrier Penetration ■ Category 1: BBB+; Category 0: BBB-; ■ The output value is the probability of being BBB+
Fu	76.24	🟢	■ The fraction unbound in plasms ■ Low: <5%; Middle: 5~20%; High: > 20%
OATP1B1 inhibitor	0.969	🔴	■ Category 0: Non-inhibitor; Category 1: inhibitor. ■ The output value is the probability of being inhibitor, within the range of 0 to 1.
OATP1B3 inhibitor	0.983	🔴	■ Category 0: Non-inhibitor; Category 1: inhibitor. ■ The output value is the probability of being inhibitor, within the range of 0 to 1.
BCRP inhibitor	0.129	🟢	■ Category 0: Non-inhibitor; Category 1: inhibitor. ■ The output value is the probability of being inhibitor, within the range of 0 to 1.
MRP1 inhibitor	0.066	🟢	■ Category 0: Non-inhibitor; Category 1: inhibitor. ■ The output value is the probability of being inhibitor, within the range of 0 to 1.

5. Metabolism

Property	Value	Decision	Comment
CYP1A2 inhibitor	0.001	🟢	■ Category 1: Inhibitor; Category 0: Non-inhibitor; ■ The output value is the probability of being inhibitor.

CYP1A2 substrate	0.0	●	■ Category 1: Substrate; Category 0: Non-substrate; ■ The output value is the probability of being substrate.
CYP2C19 inhibitor	0.0	●	■ Category 1: Inhibitor; Category 0: Non-inhibitor; ■ The output value is the probability of being inhibitor.
CYP2C19 substrate	0.0	●	■ Category 1: Substrate; Category 0: Non-substrate; ■ The output value is the probability of being substrate.
CYP2C9 inhibitor	0.0	●	■ Category 1: Inhibitor; Category 0: Non-inhibitor; ■ The output value is the probability of being inhibitor.
CYP2C9 substrate	0.22	●	■ Category 1: Substrate; Category 0: Non-substrate; ■ The output value is the probability of being substrate.
CYP2D6 inhibitor	0.003	●	■ Category 1: Inhibitor; Category 0: Non-inhibitor; ■ The output value is the probability of being inhibitor.
CYP2D6 substrate	0.0	●	■ Category 1: Substrate; Category 0: Non-substrate; ■ The output value is the probability of being substrate.
CYP3A4 inhibitor	0.0	●	■ Category 1: Inhibitor; Category 0: Non-inhibitor; ■ The output value is the probability of being inhibitor.
CYP3A4 substrate	0.0	●	■ Category 1: Substrate; Category 0: Non-substrate; ■ The output value is the probability of being substrate.
CYP2B6 inhibitor	0.0	●	■ Category 1: Inhibitor; Category 0: Non-inhibitor; ■ The output value is the probability of being inhibitor.
CYP2B6 substrate	0.0	●	■ Category 1: Substrate; Category 0: Non-substrate; ■ The output value is the probability of being substrate.
CYP2C8 inhibitor	0.357	●	■ Category 1: Inhibitor; Category 0: Non-inhibitor; ■ The output value is the probability of being inhibitor.
HLM Stability	0.005	●	■ human liver microsomal (HLM) stability ■ Category 0: stable+ (HLM > 30 min); Category 1: unstable- (HLM ≤ 30 min). The output value is the probability of human liver microsomal instability, where a value closer to 1 indicates a higher likelihood of instability. The range is between 0 and 1.

6. Excretion

Property	Value	Decision	Comment

CL_{plasma}	2.063	●	■ The unit of predicted CLplasma penetration is ml/min/kg. >15 ml/min/kg: high clearance; 5-15 ml/min/kg: moderate clearance; < 5 ml/min/kg: low clearance.
$T_{1/2}$	2.138	●	■ The unit of predicted T1/2 is hours. ■ ultra-short half-life drugs: 1/2 < 1 hour; short half-life drugs: T1/2 between 1-4 hours; intermediate short half-life drugs: T1/2 between 4-8 hours; long half-life drugs: T1/2 > 8 hours.

7. Toxicity

Property	Value	Decision	Comment
hERG Blockers	0.017	●	■ Molecules with IC50 ≤10μM or ≥50% inhibition at 10 μM were classified as hERG+ (Category 1), ■ while molecules with IC50 >10μM or < 50% inhibition at 10μM were classified as hERG - (Category 0). ■ The output value is the probability of being hERG+, within the range of 0 to 1.
hERG Blockers (10um)	0.238	●	■ Molecules with IC50 ≤10 μM are classified as hERG+ (Category 1), ■ and molecules with IC50 > 10μM are classified as hERG- (Category 0). ■ The output value is the probability of being hERG+, within the range of 0 to 1.
DILI	0.456	●	■ Drug Induced Liver Injury. ■ Category 1: drugs with a high risk of DILI; ■ Category 0: drugs with no risk of DILI. ■ The output value is the probability of being toxic.
AMES Mutagenicity	0.776	●	■ AMES Toxicity ■ Category 1: Ames positive(+); ■ Category 0: Ames negative(-); ■ The output value is the probability of being toxic.
Rat Oral Acute Toxicity	0.092	●	■ Rat Oral Acute Toxicity. ■ Category 0: low-toxicity, > 500 mg/kg; ■ Category 1: high-toxicity; < 500 mg/kg. ■ The output value is the probability of being toxic, within the range of 0 to 1.
FDAMDD	0.023	●	■ FDA Maximum (Recommended) Daily Dose. ■ Category 1: FDAMDD (+); ■ Category 0: FDAMDD (-); The output value is the probability of being positive.
Skin Sensitization	0.931	●	■ Category 1: Sensitizer; ■ Category 0: Non-sensitizer. ■ The output value is the probability of being toxic, within the range of 0 to 1.
Carcinogenicity	0.196	●	■ Category 1: carcinogens; ■ Category 0: non-carcinogens; ■ The output value is the probability of being toxic.

Eye Corrosion	0.034	●	■ Eye Corrosion ■ Category 1: corrosives; Category 0: noncorrosives; ■ The output value is the probability of being corrosives.
Eye Irritation	0.945	●	■ Eye Irritation ■ Category 1: irritants; Category 0: nonirritants; ■ The output value is the probability of being irritants.
Respiratory	0.052	●	■ Category 1: respiratory toxicants; ■ Category 0: non-respiratory toxicants. ■ The output value is the probability of being toxic, within the range of 0 to 1.
Human Hepatotoxicity	0.485	●	■ Human Hepatotoxicity ■ Category 1: H-HT positive(+); ■ Category 0: H-HT negative(-); ■ The output value is the probability of being toxic.
Drug-induced Nephrotoxicity	0.591	●	■ Category 0: non-nephrotoxic (-); ■ Category 1: nephrotoxic (+). ■ The output value is the probability of being nephrotoxic (+), within the range of 0 to 1.
Ototoxicity	0.623	●	■ Category 0: non-ototoxicity (-); ■ Category 1: ototoxicity (+). ■ The output value is the probability of being ototoxicity (+), within the range of 0 to 1.
Hematotoxicity	0.199	●	■ Category 0: non-hematotoxicity (-); ■ Category 1: hematotoxicity (+). ■ The output value is the probability of being hematotoxicity (+), within the range of 0 to 1.
Genotoxicity	0.146	●	■ Category 0: non-Genotoxicity (-); ■ Category 1: Genotoxicity (+). ■ The output value is the probability of being ototoxicity (+), within the range of 0 to 1.
RPMI-8226 Immunitoxicity	0.081	●	■ Category 0: non-cytotoxicity (-); ■ Category 1: cytotoxicity (+). ■ The output value is the probability of being ototoxicity (+), within the range of 0 to 1.
A549 Cytotoxicity	0.064	●	■ Category 0: non-cytotoxicity (-); ■ Category 1: cytotoxicity (+). ■ The output value is the probability of being ototoxicity (+), within the range of 0 to 1.
Hek293 Cytotoxicity	0.039	●	■ Category 0: non-cytotoxicity (-); ■ Category 1: cytotoxicity (+). ■ The output value is the probability of being ototoxicity (+), within the range of 0 to 1.
Drug-induced Neurotoxicity	0.063	●	■ Category 0: non-neurotoxic (-); ■ Category 1: neurotoxic (+). ■ The output value is the probability of being neurotoxic (+), within the range of 0 to 1.

8. Environmental toxicity

Property	Value	Comment
Bioconcentration Factors	0.131	■ Bioconcentration factors are used for considering secondary poisoning potential and assessing risks to human health via the food chain. ■ The unit is –log10[(mg/L)/(1000*MW)]
IGC_{50}	1.269	■ Tetrahymena pyriformis 50 percent growth inhibition concentration. ■ The unit is –log10[(mg/L)/(1000*MW)]
$LC_{50}FM$	1.481	■ 96-hour fathead minnow 50 percent lethal concentration. ■ The unit is –log10[(mg/L)/(1000*MW)]
$LC_{50}DM$	2.219	■ 48-hour daphnia magna 50 percent lethal concentration. ■ The unit is –log10[(mg/L)/(1000*MW)]

9. Tox21 pathway

Property	Value	Decision	Comment
NR-AhR	0.0	●	■ Aryl hydrocarbon receptor ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
NR-AR	0.001	●	■ Androgen receptor ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
NR-AR-LBD	0.0	●	■ Androgen receptor ligand-binding domain ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
NR-Aromatase	0.0	●	■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
NR-ER	0.028	●	■ Estrogen receptor ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
NR-ER-LBD	0.0	●	■ Estrogen receptor ligand-binding domain ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
NR-PPAR-gamma	0.0	●	■ Peroxisome proliferator-activated receptor gamma ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
SR-ARE	0.004	●	■ Antioxidant response element ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.

SR-ATAD5	0.0	●	■ ATPase family AAA domain-containing protein 5 ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
SR-HSE	0.0	●	■ Heat shock factor response element ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
SR-MMP	0.0	●	■ Mitochondrial membrane potential ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.
SR-p53	0.0	●	■ p53, a tumor suppressor protein ■ Category 1: actives ; ■ Category 0: inactives; ■ The output value is the probability of being active.

10. Toxicophore Rules

Property	Value	Comment
Acute Toxicity Rule	0	■ 20 substructures; ■ acute toxicity during oral administration
Genotoxic Carcinogenicity Rule	0	■ 117 substructures; ■ carcinogenicity or mutagenicity
NonGenotoxic Carcinogenicity Rule	0	■ 23 substructures; ■ carcinogenicity through nongenotoxic mechanisms
Skin Sensitization Rule	1 alerts	■ 155 substructures; ■ skin irritation
Aquatic Toxicity Rule	2 alerts	■ 99 substructures; ■ toxicity to liquid(water)
NonBiodegradable Rule	1 alerts	■ 19 substructures; ■ non-biodegradable
SureChEMBL Rule	0	■ 164 substructures; ■ MedChem unfriendly status
FAF-Drugs4 Rule	0	154 toxic substructures from FAF-Drug4

References

[1] Open Systems Pharmacology. Atazanavir PBPK Model Evaluation Report. 2024. Available from: https://github.com/Open-Systems-Pharmacology/OSP-PBPK-Model-Library/blob/master/Atazanavir/Atazanavir_evaluation_report.pdf. Used as a reference for Atazanavir's pharmacokinetic modeling and comparison in this study.

[2] Aulton ME, Taylor KMG. Aulton's Pharmaceutics: The Design and Manufacture of Medicines. Referenced for MCC's fraction unbound (fu) and its inert, non-ionizable nature.

[3] Benet LZ, Hosey CM, Ursu O, Oprea TI. BDDCS, the Rule of 5 and Drugability. *Advanced Drug Delivery Reviews*. 2016;101:89–98. Available from: https://doi.org/10.1016/j.addr.2016.05.007. Used to discuss the significance of Lipinski's Rule of Five in evaluating YCT-529's oral bioavailability.

[4] Bienfait B, Ertl P. JSME: A free molecule editor in JavaScript. *Journal of Cheminformatics*. 2013;5(1):24. Available from: https://doi.org/10.1186/1758-2946-5-24. Applied to explain the JSME editor's technical implementation, emphasizing its ability to generate SMILES strings and its compatibility with cheminformatics systems.

[5] BioGears Documentation. Available at: `https://www.biogearsengine.com/documentation/index.html`. Accessed December 12, 2024.

[6] Cheminfo.org. SMILES generator and checker. Retrieved December 8, 2024, from Cheminfo.org. This tool was referenced to explain how the SMILES string for YCT-529 was generated using the JSME molecule editor integrated into the Cheminfo.org website.

[7] Edginton AN, Thelen K, Willmann S. Physiologically based pharmacokinetics modeling for pediatric drug development. *Journal of Clinical Pharmacology*. 2008;48(9):1083–1097. Available from: https://doi.org/10.1177/0091270008320075. This reference was used to explain how physiologically based pharmacokinetic (PBPK) models, such as those implemented in PK-Sim, can simulate drug behavior in the absence of clinical data and guide future experimental studies. Also used to explain population-based modeling capabilities within PK-Sim.

[8] Frontiers Partnerships. Racial differences in tacrolimus pharmacokinetics: CYP3A5 polymorphisms and dose adjustments. *Transplant International*, 2024. https://www.frontierspartnerships.org/journals/transplant-international/articles/10.3389/ti.2024.13495/full. This article was referenced to

provide context on how CYP3A5 polymorphisms influence tacrolimus dosing requirements across racial groups.

[9] Gabrielsson J, Weiner D. Pharmacokinetic and Pharmacodynamic Data Analysis: Concepts and Applications. CRC Press. 2006. This book was referenced to explain the relationship between Cp, Vd, CL, and their impact on plasma concentration profiles.

[10] Gleeson MP. Generation of a Set of Simple, Interpretable ADMET Rules of Thumb. *Journal of Medicinal Chemistry*. 2008;51(4):817–834. Available from: https://doi.org/10.1021/jm701122q. Referenced to explain the implications of the GSK rule and its rejection for YCT-529.

[11] Grogan S, Preuss CV. Pharmacokinetics. In StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023. This paper was used to define and explain the importance of ADME processes in pharmacokinetics. Was also used to define key pharmacokinetic parameters, such as Cmax, Tmax, and AUC, relevant to the study objectives. Used for insights on ADME processes, particularly protein binding and bioavailability implications.

[12] Heine RT, Hillebrand MJ, Rosing H, van Gorp EC, Mulder JW, Beijnen JH, Huitema AD. Identification and profiling of circulating metabolites of atazanavir, an HIV protease inhibitor. Drug Metab Dispos. 2009 Sep;37(9):1826-40. doi: 10.1124/dmd.109.028258. Available from: http://dmd.aspetjournals.org.

[13] Julier, P. W., Zhang, L., & Enck, P. The placebo effect: Advances in understanding and clinical implications. *Trends in Pharmacological Sciences*, 2021. https://doi.org/10.1016/j.tips.2021.05.006. This article was referenced to contextualize the limitations of MCC as a placebo in pharmacokinetic simulations and its potential impact on drug absorption dynamics.

[14] Kimmelman J, Mogil JS, Dirnagl U. Distinguishing between exploratory and confirmatory preclinical research will improve translation. *PLOS Biology*. 2014;12(5):e1001863. Available from: https://doi.org/10.1371/journal.pbio.1001863. This paper was cited to highlight the importance of exploratory research in generating robust hypotheses and addressing knowledge gaps when empirical data are unavailable, emphasizing its role in drug development.

[15] Li F, Shi S, Yi J, et al. ADMETlab 3.0: An updated comprehensive online ADMET prediction platform enhanced with broader coverage, improved performance, API functionality, and decision support. *Nucleic Acids Research*. 2024;52(W1):W422–W431. Available from: https://doi.org/10.1093/nar/gkae236.

[16] Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews*. 2001;46(1–3):3–26. Available from: https://doi.org/10.1016/S0169-409X(00)00129-0. Explained Lipinski's Rule of Five and its relevance for orally administered drugs.

[17] Litherland NB, Thivierge MC. Basics of Cellulose Chemistry. *Journal of Dairy Science*. 2008. Referenced for MCC's chemical composition and neutral behavior.

[18] Mamada H, Iwamoto K, Nomura Y, Uesawa Y. Predicting blood-to-plasma concentration ratios of drugs from chemical structures and volumes of distribution in humans. *Journal of Pharmaceutical Health Care and Sciences*. 2021;47(1):1–12. Available from: https://doi.org/10.1007/s11030-021-10186-y.

[19] Mansoor A, Mahabadi N. Volume of Distribution. In StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2023. This was used to describe the significance of blood-plasma concentration and distribution models, explain the impact of lipophilicity on YCT-529's volume of distribution and plasma concentration. Provides information on distribution models and the significance of high lipophilicity in multi-compartment distribution.

[20] National Center for Biotechnology Information (NCBI). PubChem Compound Summary for CID 162679554, YCT529 free acid. Retrieved September 14, 2024, from https://pubchem.ncbi.nlm.nih.gov/compound/YCT529-free-acid. This source was used to retrieve chemical and physical property data for YCT-529, including its molecular structure, which informed the SMILES generation and subsequent pharmacokinetic modeling.

[21] Niederreither, K., & Dollé, P. Retinoic acid in development: towards an integrated view. *Nature Reviews Genetics*, 9(7), 541–553. 2008. https://doi.org/10.1038/nrg2340. I make use of this paper in providing evidence to back up RAR, RXR, and their binding with RA-related claims in my paper. It provides a comprehensive review of the role of retinoic acid in development.

[22] Niederalt C, Kuepfer L, Solodenko J, Eissing T, Siegmund HU, Block M, Willmann S, Lippert J. A generic whole-body physiologically based pharmacokinetic model for therapeutic proteins in PK-Sim. *J Pharmacokinet Pharmacodyn*. 2018 Apr;45(2):235–257. doi: 10.1007/s10928-017-9559-4. PMID: 29234936; PMCID: PMC5845054. Used to support the predictive capabilities of PK-Sim in modeling lipophilic drugs and validating its applicability to YCT-529.

[23] Noman MAA, Kyzer JL, Chung SSW, Wolgemuth DJ. Retinoic acid receptor antagonists for male contraception: current status. *Biology of Reproduction*. 2020;103(2):390–399. Referenced for insights into YCT-529's metabolic pathways.

[24] Ogwuche P. Advancing male contraception: Reviewing existing methods and the innovative approach of YCT-529 through retinoic acid receptor-α inhibition [Capstone paper]. Minerva University. 2024.

[25] Price G, Patel DA. Drug Bioavailability. In: *StatPearls [Internet]*. Treasure Island (FL): StatPearls Publishing; 2023. This was referenced to explain bioavailability and its relevance to YCT-529's clinical success, highlighting the importance of solubility and bioavailability in affecting YCT-529's therapeutic profile. Also explains bioavailability and the impact of lipophilicity on plasma drug concentration, which is crucial for YCT-529's pharmacokinetics.

[26] PubMed. Ethnic variability in warfarin dosing: Implications for pharmacogenomic-guided therapy. *Clinical Pharmacology Therapeutics*, 2024. https://pubmed.ncbi.nlm.nih.gov/15855242/. This article was cited to illustrate how genetic differences in CYP2C9 and VKORC1 impact warfarin metabolism and dose optimization.

[27] PubMed. Pharmacogenetic variability in drug metabolism: The impact on dose-response relationships. *Annual Review of Pharmacology and Toxicology*, 2005. https://pubmed.ncbi.nlm.nih.gov/16984210/. This reference was used to support the broader discussion of pharmacogenomic influences on interindividual drug response.

[28] Roberts, C.J., et al. (2015). Role of transporters in the disposition of drugs. AAPS J.

[29] Rowe RC, Sheskey PJ, Quinn ME. Handbook of Pharmaceutical Excipients. 2009. Used to determine LogP, solubility, and the molecular characteristics of MCC.

[30] Rowland, Y., Balant, L. P., & Aarons, L. Physiologically based pharmacokinetics: Current status and future directions. *Clinical Pharmacokinetics*, 2018. https://doi.org/10.1007/s40262-018-0694-5. This article was cited to highlight the advantages of integrating PBPK and PopPK modeling approaches to improve inter-individual variability predictions in pharmacokinetics.

[31] Shi, R., Wolgemuth, D. J., & Georg, G. I. Development of the retinoic acid receptor alpha-specific antagonist YCT-529 for male contraception: A brief review. *Contraception*, 2024. https://doi.org/10.1016/j.contraception.2024.110809. I make use of this paper to support the discussion on YCT-529's mechanism of action, pharmacokinetics, and preclinical efficacy in male contraception.

[32] Wilkinson GR. Ethnic differences in CYP enzyme expression and drug metabolism. Clin Pharmacokinet. 2004;43(9):583–606.

[33] Willmann S, Edginton AN, Kleine-Besten M, Jantratid E, Thelen K, Dressman JB. Whole-body physiologically based pharmacokinetic population modelling of oral drug administration. *Journal of Pharmacy and Pharmacology*. 2004;61(7):891–899. Discussed PK-Sim's compartmentalization capabilities.

[34] Xu Y, Dai Z, Jin Z, Zhang L. ADMETLab 2.0: An integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic Acids Research*. 2021;49(W1):W5–W14. Available from: https://doi.org/10.1093/nar/gkab255.

[35] Yang H, Lou C, Sun L, Li J, Cai Y, Wang Z, Tang Y. admetSAR 2.0: Web-service for prediction and optimization of chemical ADMET properties. *Bioinformatics*. 2019;35(6):1067–1069. Available from: https://doi.org/10.1093/bioinformatics/bty707. Referenced for describing the reliability and methodology of ADMETLab predictions.

[36] Zanger UM, Schwab M. Cytochrome P450 enzymes in drug metabolism: Regulation of gene expression, enzyme activities, and impact of genetic variation. *Pharmacology & Therapeutics*. 2013;138(1):103–141. Cited for CYP enzyme variability and their inclusion in the model.